Food Purchase Diversity Across U.S. Markets

Mark D. Jekanowski
USDA/ERS/FRED/FMB, 1800 M St. NW, Washington, DC 20036-5831.
E-mail: MarkJ@ers.usda.gov

James K. Binkley
Purdue University, 1145 Krannert, West Lafayette, IN 47907-1145.
E-mail: Binkley@agecon.purdue.edu

ABSTRACT

Growth in the number of food products and brands carried by grocery stores implies a preference for diversity. We examine the factors that affect the variety of food purchases across U.S. markets. Three measures of variety, based on market-level sales of both grocery categories and RTE cereal brands, are regressed on various population and market characteristics. We find that markets with a high proportion of low-income individuals exhibit less variety both in terms of grocery categories and breakfast cereal brands. We also find that racial diversity and average store size are important factors in explaining diversity of grocery purchases across markets, but are less important in explaining the variety of cereal brand purchases. [EconLit citations: D120, L660.] © 2000 John Wiley & Sons, Inc.

1. INTRODUCTION

Analysis of the relationship between food expenditures and income dates at least to Engel. It is well known that as income increases, the proportion of the budget devoted to food decreases. Once basic needs are met, income growth allows consumers to add items to their purchase set that might otherwise be considered discretionary, or even luxuries. Thus, the variety of products purchased is expected to increase as consumer budgets expand. In examining aggregate expenditures across countries in various stages of development, Theil and Finke (1983) have documented a positive relationship between the variety of items purchased and the level of income.

A similar phenomenon applies to subsets of goods, and certainly to food. Food expenditures increase with income, but in a developed economy, increased food expenditures are unlikely to reflect a greater quantity of each food item, for this would imply a proportionate increase in eating. Rather, the mix of items purchased will change as consumers are able to add more varied, higher quality goods, appealing to preferences rather than just to basic needs. In a study based on Stigler's (1945) famous diet problem, Silberberg (1985) tested this hypothesis. Using survey data, he calculated average nutrition levels achieved by income groups, then calculated the minimum expenditures required to achieve these levels. He found that the difference between actual and minimum expenditures increased with income, and concluded that increasing food expenditures with higher income are to please tastes rather to meet nutritional needs. Jackson (1984) demonstrated that this result holds

Agribusiness, Vol. 16, No. 4, 417–433 (2000)

across many subsets of goods, including food, housing, recreation, and several others. Other studies, often using survey data, have shown that the number of different food items purchased by households tends to increase as food expenditures increase (Lee & Brown, 1989; Lee, 1987; Shonkwiler, Lee, & Taylor, 1987). These also illustrate the role of demographic characteristics, such as age, income class, and household size.

The variety of consumer products available in the grocery industry is especially large, and continues to grow. Each year, on average, more than 12,000 new grocery items are introduced. While some of these can be considered new products, most can be considered extensions of existing items and brands, with additions of new flavors, package sizes, or cooking alternatives (Gallo, 1999). Between 80% and 95% of these eventually fail, often replaced by yet another variant of a similar item. Some categories, such as ready-to-eat (RTE) breakfast cereal, have even been the focus of studies alleging that brand proliferation is primarily an anticompetitive strategy that reduces social welfare (e.g., Scherer, 1979; Connor, 1981).

Expansion in categories, items, and brands has been facilitated by trends in the retail industry. The 1987 annual report of *Progressive Grocer* declared that "the supermarket industry is moving faster to accommodate changes in consumer shopping and eating patterns." While the total number of supermarkets in the U.S. declined from about 26,800 in 1980 to 24,500 in 1994, the number classified as "superstores" increased from about 3,200 to more than 6,500 in 1994. During this period, the average square footage of grocery sales space per store increased from about 23,000 to around 35,000, and the average number of unique brands, package sizes, and flavor items carried by supermarkets rose from 14,000 to 25,000 (Food Marketing Review), 1996. Messinger and Narasimham (1997) find this to be not an effort to lower costs through scale economies but a method to increase consumer convenience and provide "one-stop shopping." An important factor is the pressure from intrusion of combination stores by retailers such as Walmart, in grocery markets.[1] Increasing store size reflects the new store formats and the response of conventional chains to this competition. *The Wall Street Journal*, in a 1997 article on this response, notes that "chains are expanding and remodeling their stores," finding this "more effective than price cutting." (p. B11).

A second trend is the increasing percentage of consumer spending being taken by restaurants, especially fast food. The same *Progressive Grocer* report states this has "increasingly disturbed" grocery industry leaders, finding in its store survey that two thirds of supermarket managers rated the competitive threat from fast food as moderate or serious. In a study of food-away-from-home demand, Hiemstra and Kim (1995) concluded that "fast food is more of a commodity than most other food-away-from-home and it competes more strongly with food purchased at grocery stores than with other food-away-from home." (p. 30). Hence we have supermarkets offering prepared foods and more items designed for convenience.

While consumer preference for variety is well documented at the household level, the extent to which this is reflected in aggregate sales data is not. This is an important question, given the ubiquitous use of aggregate data in demand studies, and the fact that food industry managers tend to be interested in market characteristics, rather than the particular behavior of individual consumers, when developing marketing strategies. We examine the ex-

[1]Progressive Grocer (1991) defines a superstore as "A supermarket with at least 30,000 square feet, doing more than $8 million annually and offering an expanded selection of non-foods. Specialty departments and extensive services are offered." A combination store is defined as "a superstore but percentage of non-food space is 40% or more." (p. 6)

tent to which consumer preference for variety is reflected in aggregate sales shares across U.S. markets, and determine the role of market characteristics and demographics in shaping market behavior. An important aspect of the study is that we consider variety in both products (categories) and brands (within category). To the best of our knowledge, brand variety has not previously been studied at any market level.

2. THEORETICAL DEVELOPMENT

Previous studies of preference for variety rely on the theory of the consumer. For a representative consumer, given prices and income, the quantity demanded of some good(s) will be zero. This situation reflects a binding budget constraint. As income—or subgroup expenditures—rise, or prices fall, the number of items with positive demand is expected to increase, since the consumer can now afford items previously considered too expensive or discretionary. It follows that the variety of purchases increases with expenditures. As noted above, demographic characteristics of individual consumers can also affect consumer preference for variety (e.g., Shonkwiler et al., 1987; Lee, 1987).

But the number of products purchased is only part of the story. It is also important how expenditure shares are allocated among new and existing products in the market basket. Increased purchases of certain products are likely to be accompanied by decreased purchases of others, as the quantity of food consumed remains relatively constant. For example, as incomes (or expenditures) rise, the share of the food budget spent on prepared foods might increase, while the share spent on basic staple goods, such as flour, would be expected to decrease. This has been documented in several studies (e.g., Park & Capps, 1997; Nayga, 1998).

Demand equations for particular products are often expressed in share form as follows:

$$w_i = \frac{p_i q_i}{x_F} = h_{Fi}(p_i, x_F, z) \tag{1}$$

where w_i is the proportion of food expenditures spent on product i, p_i is the price of that product, x_F is subgroup expenditures, and z controls for individual consumer preferences or demographic characteristics. As a measure of the diversity of the market basket, we are interested in the distribution of w_i across all available goods in the subgroup. An individual consumer maximizes the diversity of products purchased by allocating an equally small share of total expenditures to all products in the subgroup (Patil & Taillie, 1982). Define d to be a measure of the equality of budget shares within the subgroup, i.e., diversity:

$$d = g_F(P_F, x_F, z). \tag{2}$$

Unlike equation (1), equation (2) does not specify demand for a particular product; rather, it describes how shares are distributed across all products. The relevant price is the price level of all goods in the subgroup (P_F), possibly expressed as an index. Subgroup expenditures and consumer characteristics will have independent effects on the diversity of products purchased.

Our focus is on aggregate consumer behavior across independent U.S. markets. Thus, we examine how expenditures are allocated across a subset of grocery products in each market. To the extent that expenditure levels and population characteristics such as income and demographics vary across markets, we expect market-level data to reflect the behavior of a

representative consumer. This is one hypothesis to be tested empirically. Market-level data also enable us to control for variations in market structure that could facilitate or encourage variety-seeking behavior. Of particular interest is the trend toward larger stores with wider product offerings, as mentioned above. Variation in average store size across markets should proxy consumers' ability to easily choose a diverse market basket, and hence should be reflected in aggregate sales data. In addition, a market focus permits a consideration of aspects of a variety difficult to examine otherwise. For example, if households of different races purchase similar quantities but different items, race will not affect household variety, but may impact market variety. No previous studies have accounted for factors such as these.

Market data facilitates another unique aspect of our study: examination of diversity of expenditures at the brand level as well as at the product level. For products, purchase diversity across 338 packaged grocery categories—including frozen foods and refrigerated items—is examined. At the brand level, the focus is on the ready-to-eat (RTE) breakfast cereal segment, consisting of over 160 individual brands. This is the food product category with greatest sales, and, as noted above, has been the focus of much previous research. For both products and brands, we examine the role of subgroup expenditures and various population characteristics in explaining variations in purchase diversity.

2.1. Measuring Diversity

Measures of diversity and its dual concept, concentration, have several applications in economics. The most common use involves studies of market structure, where the distribution of market shares is often used as an indicator of market power. The Herfindahl Index, and measures based on entropy, have been used in this arena.

The diversity of a market basket comprised of n goods can be measured in a similar fashion. Expenditure shares concentrated among only a few items are an indication of little diversity. Maximum diversity occurs when expenditure shares are equally distributed among all products. Entropy and the Herfindahl Index are useful measures of the distribution of expenditure shares and have been used in this context before (Theil & Finke, 1983; Lee & Brown, 1989).

Entropy (E) is defined as:

$$E = -\sum_{i=1}^{n} w_i \log w_i \tag{3}$$

and the Herfindahl index (H) is:

$$H = \sum_{i=1}^{n} w_i^2 \tag{4}$$

where w is the budget share of commodity i.

Entropy can vary from zero (when one budget share equals 1 and hence the $n - 1$ others vanish) to a maximum of log n (when all shares equal $1/n$). The Herfindahl index measures concentration instead of diversity, varying from one, when the entire budget share is spent on 1 product, to $1/n$ when all budget shares are equal. We use $1 - H$ as a direct measure of diversity. This is often referred to as the Simpson index.

The shape of the cumulative distribution of expenditure shares can also provide information about diversity. If expenditures are uniformly distributed across all products (implying maximum diversity), the sum of the shares of any x products will simply be x/n. On the other hand, if expenditures are concentrated on particular items, the cumulative distribution of expenditure shares, based on a ranking of shares from highest to lowest, will be skewed to the left. We construct an additional measure of diversity based on this principle.

For each market we ranked the expenditure shares of the products in each subgroup from highest to lowest, and computed the sums of these shares at levels representing approximately 75% of total (grocery or category) sales. We refer to these as cumulative shares (CS). Across grocery categories, our CS measure is based on the sums of the top 75 expenditure shares, while within the breakfast cereal category the sums of the largest 50 cereal brand shares are used. When the CS is relatively low, the implication is greater purchase diversity, since the most important (i.e., largest sales) items receive a relatively low percent of total grocery spending. Since this measures concentration instead of diversity, it is multiplied by -1 so that the signs of the parameters correspond with those from the Simpson and entropy measures.

It is reasonable to expect these three measures to yield similar results. However, they tend to emphasize different aspects of diversity. Entropy places greater weight on smaller shares, so it is especially sensitive to differences in the number of minor commodities in the market basket. The cumulative share measures ascribe variety as characteristic of a basket in which there is a relatively low number of predominating goods, while saying nothing about how the shares of the minor products are distributed. The Simpson index is somewhat intermediate between these two.

3. DATA AND EMPIRICAL MODEL

All data on product sales came from Sales Area Marketing, Inc. (SAMI), a product tracking firm which discontinued operation in 1991. SAMI tracked grocery product sales by monitoring grocers' warehouse shipments in 54 market areas, areas that accounted for 85%of U.S. grocery sales. The areas were aggregations of counties around major urban centers, chosen based upon warehouse shipping patterns.[2] The warehouses provided SAMI with data on case movements to supermarkets, along with corresponding retail sales and price data, with considerable care taken to correct the data for movements into and out of the region. Four hundred eighty-four product categories were tracked. Of these, 339 were food items, the balance being pet foods, paper and cleaning products, and health and beauty aids.

SAMI mainly sold the data to manufacturers, making it available in various forms. This study used the "Market Development Indices." These indices report average household expenditures on each category, in each market, as indices relative to U.S. average household expenditures on that category. This study used annual data for 1990, the last year available. The categories covered are listed in the Appendix.

These indices are useful for examining how expenditures on individual products vary across markets. In this sample, each market has positive sales for at least 333 of the 338 grocery categories covered. Converting these indices to regional expenditure shares is straightforward, using total U.S. category sales and regional grocery expenditures.

[2]On average, an area was comprised of 35 counties, with total area population of four million.

Table 1. Descriptive Statistics for Variables Used in the Empirical Analysis

Variable	Mean	Standard Deviation	Range
Grocery Shares			
Entropy	4.893	0.027	4.83–4.96
Simpson	0.985	0.001	0.98–0.99
CR_{75}	0.736	0.009	0.71–0.75
Expenditure index	0.996	0.121	0.74–1.31
Cereal Brand Shares			
Entropy	4.394	0.059	4.23–4.49
Simpson	0.983	0.002	0.97–0.99
CR_{50}	0.730	0.021	0.67–0.76
Expenditure index	1.017	0.163	0.68–1.51
Independent Variables			
Per capita income ($1000)	13.667	2.078	9.97–18.94
Gini Coefficient of income distribution	0.420	0.017	0.39–0.46
Proportion of population in poverty	0.133	0.041	0.08–0.25
Female labor force part. rate	0.566	0.038	0.40–0.65
Proportion college graduates	0.194	0.038	0.11–0.29
Proporion of population under 14 yrs.	0.219	0.020	0.18–0.30
Proportion of population over 65 yrs.	0.126	0.022	0.08–0.19
Proportion of population that is white	0.880	0.087	0.66–0.99
Grocery store floor space (1000 sq ft)	21.633	2.975	17.21–30.81

For certain categories, SAMI reports brand shares within each market. RTE cereal, our focus, has virtually complete coverage. Reported for each market are the expenditure shares of 160 cereal brands, accounting for well over 90% of total cereal sales in each market. All cereal brands are not available in every market: on average there are approximately 154 brands available per market.

Entropy and Simpson indices of expenditure shares, along with the cumulative share measures described above, are used as dependent variables in regression equations describing the factors influencing purchase diversity. Based on 338 grocery products, maximum diversity would imply an entropy value of 5.823, and a Simpson index of 0.997. Based on an average of 154 cereal brands available per market, maximum entropy is 5.037, and the maximum Simpson index is 0.994. Across markets the variation in these measures of diversity is slight (see Table 1), which lends special importance to using more than one method to quantify diversity. Results which are consistent across different specifications of the dependent variable can increase our confidence in the results.

Independent variables include market-level measures of grocery or breakfast cereal expenditures, demographic characteristics of the market population, and a measure of average store size in each of the SAMI markets. Grocery and breakfast cereal expenditures come from SAMI. For the analysis of grocery product purchases, an index of per capita food store sales across regions is used (per capita food store sales in each region relative to the U.S. average). Within the breakfast cereal category, the market development index for breakfast cereals is used as a measure of regional category household sales. We expect diversity of purchases to be positively related to its respective measure of total subgroup expenditures, a relationship that has been found by previous studies based on household data.

Market-level demographic characteristics were compiled by aggregating 1990 U.S. census county data to correspond to each SAMI market area. Household characteristics include two age categories (the proportion of the population between 1 and 14 years, and the proportion over 65 years), the proportion of the population that is white (an inverse measure of ethnicity), percent of the population that has at least a 4-year college degree, and the percent of households with income below the poverty line.

We also include per capita income and the female labor force participation rate, two variables which can be interpreted as measures of consumer time costs within each market. Time costs could affect purchase behavior either through shopping costs or by changing the types of products purchased on the basis of the amount of home preparation required. Female labor force participation can affect time costs directly by decreasing the amount of time available for both activities. Per capita income is likely correlated with the wage rate, and high incomes also often involve more time working, increasing the opportunity cost of time.[3] But income class might also affect the types of foods purchased due to tastes or customs. To account for differences in consumption behavior due to income class, we also include the Gini coefficient of income distribution within each market area. This was constructed from county-level data based on the proportion of the population in each of 14 income classes.[4] The store size variable, for which we expect a positive effect on diversity, was constructed from Progressive Grocer 1991 state measures based on SAMI market populations in the respective states. Descriptive statistics of all variables used in the analysis, including the dependent variables, are presented in Table 1.

Three indicator variables (for the South, West, and Midwest regions; Northeast region omitted) were included in each model to control for regional variations in tastes and possible price differences. If the general price level for all grocery products (or all cereal products) is lower in any of the four regions, greater diversity would be expected since more products can be purchased for any given level of total expenditures.

3.1. Model Estimation

To estimate the effect of the independent variables on diversity of purchases, the following linear model was estimated:

$$d = f(\text{subgroup expenditures, population characteristics, market characteristics})$$

where d is purchase diversity measured each of three ways: Simpson index of expenditure shares, entropy of expenditure shares, and a cumulative share (CS) measure. This relationship was estimated both across grocery categories, and across brands within the breakfast cereal category.

The models explaining the Simpson and entropy indices were estimated using Ordinary Least Squares (OLS). Since the CS measures represent proportions, their variation is best explained using a minimum χ^2 type estimator instead of OLS. We assume that CS_i =

[3]Also, higher incomes permit more attractive uses of time not working, making shopping and cooking "inferior" goods.

[4]The Gini coefficient is a numerical measure of income inequality ranging between 0 (for perfect equality) and 1 (for absolute inequality). Calculation is based on the proportion of total income earned by each culmulative percentage of the total population, i.e. if a "small" proportion of the population earns a "large" proportion of national income, the Gini coefficient will approach 1. For the U.S. as a whole, most studies report a Gini coefficient of about 0.4.

$\Phi(\beta' x_i)$, where $\Phi(\bullet)$ is the distribution function of the standard normal. Hence, the model we estimate is:

$$\Phi^{-1}(CS_i) = \alpha + \beta x_i + \epsilon \quad (5)$$

where $\Phi^{-1}(CS_i)$ is the probit (i.e., normal equivalent deviate) of the cumulative share in each market i, and x_i are the values of the independent variables.[5] This model was estimated using weighted least squares, with weights based on total sales quantity across markets (see Greene, 1990).

4. RESULTS

The regression results for variety across grocery categories are presented in Table 2, and the results based on breakfast cereal brands are in Table 3. Results from the Simpson and entropy measures are labeled as such, and the CS measures are labeled CS_x, representing the sum of the expenditure shares for the top x products.

4.1. Grocery Purchases

We focus first on the diversity of expenditures across the 338 grocery categories. Our results are presented in Table 2.

The results indicate that the distribution of market-level expenditures is related to demographic and other characteristics of the market in ways that are consistent with previous research based on household data. Importantly, we find that in markets where average grocery expenditures are high, expenditures are more evenly distributed across all products—implying greater diversity. This likely reflects a shift toward more discretionary items as expenditures increase, a result which is expected based on consumer theory.

The proportion of the population in poverty is significant in each equation, showing that the diversity of expenditures decreases as the proportion of low-income consumers in the market increases. This is consistent with a need for financially strapped consumers to economize by restricting purchases to basic, relatively inexpensive staple items.

Per capita income is only strongly significant in the equation predicting the cumulative share measure CS_{75}, where it reduces diversity. Much of the income effect on grocery expenditure diversity is likely to operate through food expenditures and the proportion of the population in poverty. After controlling for these factors, income has only a modest effect on diversity in different markets. However, the general direction of the effects support an interpretation of income as a time measure. The Gini coefficient, which measures how evenly income is distributed within each market, is not significant in any of the equations.

The female labor force participation rate is a more direct measure of time costs, accounting directly for the opportunity cost of shopping and home meal preparation, since female labor force participation generally implies that both spouses are employed outside the home. Results show all coefficients are negative, with Simpson the strongest, implying less variety. That entropy and CS significance are virtually the same suggests a balanced effect: less spending on minor categories coupled with more spending on a few major categories. This is consistent with a shift towards food away from home or other convenience foods

[5]Alternatively, the dependent variable could be specified as a logit, i.e., $ln(CS_i/(1\text{-}CS_i))$. In practice, the choice between a probit or logit specification has little effect on the estimates.

Table 2. Estimated Model Predicitng Purchase Variety Across Grocery Categories, Based on Simpson, Entropy, and Cumulative Share Measures[1,2]

Variable	Simpson	Entropy	CS_{75} (probit)
Intercept	99.144**	5.039**	−0.154
	(175.63)	(21.10)	(−0.85)
Grocery store expenditure index	0.252**	0.084**	0.024
	(3.06)	(2.42)	(0.83)
Per capita income ($1000)	0.002	−0.006	−0.011**
	(0.18)	(−1.21)	(−2.85)
Percent of population in poverty	−0.012*	−0.009**	−0.898**
	(−1.98)	(−3.67)	(−4.33)
Female labor force participation rate	−0.009**	−0.003	−0.255
	(−2.05)	(−1.59)	(−1.57)
Proportion college graduates	0.004	−0.001	−0.054
	(0.62)	(−0.41)	(−0.28)
Proportion of population under 14 yrs.	−0.226	0.077	−0.332
	(−0.34)	(0.27)	(−1.27)
Proportion of population over 65 yrs.	−1.461**	−0.512*	−0.657**
	(−2.39)	(−1.98)	(−3.10)
Proportion of population that is white	−0.455**	−0.156**	−0.118**
	(−3.76)	(−3.05)	(−2.77)
Gini coefficient	0.632	0.567	0.082
	(0.54)	(1.15)	(0.24)
Store size	0.012**	0.005**	0.006**
	(3.06)	(3.34)	(5.13)
Regional indicators	$F^3_{40} = 2.28$*	$F^3_{40} = 0.32$	$F^3_{40} = 3.83$**
South	0.005	−0.002	−0.003
	(0.13)	(−0.10)	(−0.27)
West	−0.065*	−0.003	0.007
	(−1.87)	(−0.21)	(0.68)
Midwest	−0.050*	0.006	0.024**
	(−1.73)	(0.50)	(2.51)
R-square	0.78	0.59	0.76

[1]T-statistics in parentheses.

[2]Single and double asterisks denotes significance at the 10% and 5% levels, respectively.

as females enter the workforce (Yen, 1993; Byrne, Capps, & Saha, 1996; McCracken & Brandt, 1987). Instead of dividing expenditures among many inputs to produce a meal at home, these consumers will tend to purchase products that are fully prepared, such as frozen entrees, or other items requiring minimal assembly at home. By purchasing fewer ingredients, hence fewer items, the result is a less diverse basket of purchases.[6]

As expected, the percent of the population that is white is negative and significant in each equation, implying that less ethnic diversity in the marketplace corresponds with less diversity in the market basket of goods purchased. It is notable that food manufacturers and marketers, perhaps belatedly, are recognizing ethnic variation in food preferences, and are creating products that are targeted toward various ethnic groups.

Results indicate that markets with a large proportion of consumers over the age of 65 tend to exhibit less diverse aggregate food expenditure bundles. Perhaps these consumers are

[6]We examined correlations between female labor force participation and food spending and found high positive correlations with frozen items and large negative correlations with items used for food preparation, especially baking ingredients.

more "set in their ways," and therefore less likely to experiment with alternative and new food products. We thus find that the two most important demographic trends in the US—rising ethnic diversity and the aging of the population—are estimated to have countervailing effects on variety.

Neither kids nor percent of the population with a college degree was found to have a significant effect on diversity of grocery purchases. The only indication of an effect is the modest significance of kids in the CS equation. This suggests somewhat more emphasis on major categories in markets with many children. That education has no impact is somewhat surprising, but it is consistent and statistically convincing across all measures.

The trend toward larger stores characterizing U.S. food retailing obviously permits the handling of a greater variety of products. Any effect of this on purchase diversity would surely be positive. Our results support this, and also suggest it is quite strong. Like increased ethnic diversity, this is another important trend, giving food manufacturers an incentive to widen the diversity of their offerings.

The regional indicator variables (Northeast region omitted) provide only limited evidence that, after accounting for the above factors, only modest regional diversity remains. F-tests of these indicators are significant at standard levels only in the Simpson and CS specifications, but the individual effects provide little evidence that any particular region has unusually high or low diversity.

4.2. Breakfast Cereal Purchases

The results for the brand-level on RTE cereals are presented in Table 3. The independent variables are the same as in our analysis of categories, with grocery expenditures replaced by household breakfast cereal expenditures. We use the same three measures, with the CS measure the sum of the expenditure shares for the top 50 cereal brands.

We find the breakfast cereal expenditures index to be significant only in the CS specification. The positive effect indicates that expenditures on the cereal brands that normally receive the highest share decrease as expenditures increase. The highest share cereals tend to be the most basic brands, such as Cheerios®, Corn Flakes, Wheaties®, and the like. These also tend to be the least expensive. So, the CS equation suggests lower expenditures on the basic, relatively inexpensive brands as expenditures increase. This is the expected effect, since greater category expenditures should imply a shift toward more expensive brands. However, as noted above, the CS measure says nothing about how minor shares are distributed. For this, we examine the entropy equation, since entropy is especially sensitive to changes in minor shares. Here, the expenditure index is not significant, implying no off-setting increase in expenditures on the most minor brands as expenditures on the largest share brands decline. A likely explanation is that as category expenditures increase, consumers shift from the basic brands to the "middle tier" brands—from brands like Grape-Nuts® to Blueberry Morning®, perhaps—leaving the extremely minor brands substantially unaffected. This differs from results at the category level, where rising variety was found to be due mainly to expansion of sales in minor categories.

Higher income is again associated with less variety, but also with a different pattern. Specifically, for cereals, the Simpson measure has the greatest statistical strength and CS the least. For categories, CS was highly significant, while for Simpson, significance was completely absent. The suggestion is that for categories, higher income is associated with greater spending on those that absorb the largest expenditures, while for cereals the effect is more generalized. Whether this is meaningful is not clear, but both support the value of time role of income and thus our initial hypothesis.

Table 3. Estimated Model Predicitng Purchase Variety Across Cereal Brands, Based on Simpson, Entropy, and Cumulative Share Measures[1,2]

Variable	Simpson	Entropy	CS_{50} (probit)
Intercept	101.077**	4.646**	0.152
	(71.77)	(10.89)	(0.32)
Breakfast cereal expenditure index	0.213	0.033	0.135**
	(1.44)	(0.74)	(2.39)
Per capita incom ($1000)	−0.056*	−0.014	−0.013
	(−1.98)	(−1.60)	(−1.35)
Percent of population in poverty	−2.950*	−0.829*	−0.316
	(−1.98)	(−1.84)	(−0.60)
Female labor force participation rate	−2.260**	−0.331	−0.754*
	(−2.01)	(−0.97)	(−1.80)
Proportion college graduates	2.557*	0.574	1.206**
	(1.66)	(1.23)	(2.35)
Proportion of population under 14 yrs.	−1.364	−0.096	−0.493
	(−0.71)	(−0.17)	(−0.68)
Proportion of population over 65 yrs.	−0.399	0.146	0.702
	(−0.25)	(0.30)	(1.15)
Proportion of population that is white	0.268	0.132	0.001
	(0.92)	(1.49)	(0.01)
Gini coefficient	−2.97	−0.365	−1.242
	(−1.04)	(−0.42)	(−1.35)
Selling area	0.010	0.005*	−0.0004
	(1.03)	(1.77)	(−0.14)
Regional indicators	$F^3_{40} = 5.33$**	$F^3_{40} = 3.85$**	$F^3_{40} = 9.35$**
South	0.042	−0.007	0.079**
	(0.44)	(−0.24)	(2.48)
West	0.270**	0.026	0.116**
	(3.47)	(1.11)	(5.09)
Midwest	0.144**	0.045**	0.041*
	(2.32)	(2.37)	(1.91)
R-square	0.78	0.73	0.72

[1]T-statistics in parentheses.
[2]Single and double asterisks denotes significance at the 10% and 5% levels, respectively.

Our other measure of time costs, the female labor force participation rate, also suggests less diversity of brand purchases in markets where consumers are time pressed. This effect is significant in the Simpson equation, and especially the CS equation. As with income, this suggests that extensive product proliferation does not provide utility to all consumers—those with the greatest time costs might prefer more streamlined product offerings. Interestingly, new product introductions in the breakfast cereal category have begun to decline in recent years (Gallo, 1999).

However, other consumers evidently desire extended choice. The variable measuring education, the percent of college graduates in each market, is significant at standard levels in two of the three models, with signs suggesting that *ceteris paribus* demand for brand variety increases with education. Highly educated consumers might be more discerning in choosing brands to closely match their tastes, so even slight variations in product characteristics will capture additional sales. These consumers likely welcome product proliferation, since each product variant introduced has a chance to more closely match the idiosyncratic preferences of any particular consumer.

The poverty variable is moderately significant only in the entropy and Simpson equations, where it indicates that low-income consumers concentrate purchases on fewer brands. Since cereal prices tend to vary considerably across brands, we expect low-income consumers to puchase primarily lower priced brands—a relationship found by Jones and Mustiful (1996). As noted above, those cereals include such brands as Cheerios®, Corn Flakes and other basic, relatively undifferentiated brands, which also tend to be the brands with the largest sales share. We note that the Gini coefficient of income distribution is not significant in any of the models.

Surprisingly, neither age variable is found to be of any significance, for any measure. This was not our initial expectation, given that many brands are believed to be specifically marketed toward particular age categories, especially children. We note that among cereals with the largest national shares are several classified as "sweetened," which is usually used to indicate children's cereals. Our results might indicate that many of these appeal to adults also, an appeal not discouraged by several notable advertising campaigns. Given that children also commonly eat "adult" cereals (perhaps by parental fiat), the result is little systematic variation in the variety of cereal purchases, on average, between children and adults.

The proportion of the population that is white, an inverse measure of ethnicity, shows little evidence of an effect on brand-share diversity. We had no strong expectation for the sign of this variable, since most brands of cereal tend to be broadly targeted across all ethnic groups. We see no inconsistency with its importance in the category model. Race and cultural differences are more likely to affect whether RTE cereal is bought in the first place rather than the particular kind selected if it is bought.

Finally, the store size variable shows only moderate effects in the brand-share models. The strongest effect is for the entropy measure, showing more variety as store size increases. Coupled with the absence of an effect for CS, the implication is that to the extent that larger stores lead to more variety in brands, it is to provide a venue for minor products. Most supermarkets carry the major brands of the categories they handle, so this outcome is not especially surprising.

F-tests of the regional indicators are highly significant in each model (with Northeast omitted), and the individual parameter estimates suggest greater cereal purchase diversity in the West and Midwest regions, which appear to be relatively homogeneous after accounting for the factors in the model.

Our results suggest that the factors examined are generally more important in explaining category variety than brand variety. Furthermore, while some factors, such as income and measures of time, affect variety at both the category and brand level, there are important exceptions. One is education, which is found to affect only brand choice. Racial diversity in the market, the age distribution, and store size are all found very important for categories, but virtually inconsequential for cereal brands. Only two factors—market income distribution and market percentage of children—were estimated to have no influence across all models. The absence of a kids effect suggests that the importance of target marketing of food products to children may be somewhat overrated.

5. CONCLUDING REMARKS

Previous studies, using household survey data, have shown that consumers prefer variety, and have related this preference to several factors. In this study, we have examined market-level data and found that many factors identified in prior studies as affecting household demand for diversity also explain differences across large market areas. We measured diver-

sity in three ways, and in general all the methods gave similar results, especially for those variables that we believe to be most important. As in previous work, we examined diversity in products, but we also were able to consider brand diversity for a single product: breakfast cereal. It was found that consumers' demand for diversity can depend on whether it is measured at the product level or the brand level.

This research emphasizes the fact that consumer utility depends on the types of products purchased, and by examining the distribution of expenditure shares across all available products, we can examine how rational consumers change their purchasing habits to reflect differences in market situations or consumer characteristics. Importantly, we find that the composition of the market basket is not static, but can be affected by many non-price factors. This is an issue of continuing interest not only to academics, but also to policy makers and those in the food industry, and it is well worth additional study.

APPENDIX I: GROCERY CATEGORIES INCLUDED IN SAMI DATA

cereal baby food
juice baby food
formula baby food
strained baby food
junior baby food
misc baby food
dessert baking mixes
piecrust mix
biscuit mix
muffin bread & roll mix
pancake mix
baking chocolate & bits
cake decorations
baking extracts
baking powder & soda
cream of tartar
corn starch
dry yeast
coconut
baking nuts
ready to spread frosting
frosting mix-double layer
frosting mix-single layer
chocolate candy bars
non-chocolate candy bars
hard roll candy-solid breath fresh
marshmallows
caramel corn
pkgd chocolate covered fruits
pkgd chocolate covered creams
pkgd chocolate covered nuts
pkgd candy covered non-chocolate
pkgd hard sugar candies
pkgd soft sugar candies
pkgd jellies
pkgd non choc chewy types
pkgd novelty items
pkgd diet candies
pkgd solid choc pieces
pkgd other chocolates
pkgd candy covered choc
ready-to-eat cereal
cereal meal bars
hot cereals
cocoa
milk modifiers
instant breakfast
coffee brewed
coffee instant
catsup
chili sauce
prepared mustard
meat sauce
Italian food sauce
barbecue sauce
tartar sauce
liquid gravy-spices & extracts
dry gravy season sauce mix
vinegar and cooking wines
misc sauces
crackers
bread & cracker crumbs
croutons
stuffing mixes
breading & batter mixes
specialty & snack crackers

ice cream cake/cups & cones
rice/grain cakes
pudding
packaged pie & fillings
maraschino cherries
gelatin desserts
gelatin-drinks-capsules
dry topping mixes
dessert & ice cream toppings
tapioca
diet sweeteners
diet canned fruit
diet canned vegetables
diet desserts
diet jams jellies & spreads
measured diet meals
low/reduced cal salad dressing
diet misc foods
canned salmon
canned sardines
canned shrimp
canned tuna
misc canned fish
corn meal
cake flour
dark flour
hominy grits
family flour
canned peaches
canned fruit cocktail
canned apple sauce
canned apricots
canned pears
canned cherries
canned mandarin oranges
canned cranberries
canned pie filling
canned pineapple
misc canned fruit
raisins
dried prunes
dates
dried apricots
dried figs
misc dried fruits
single pack gum
multiple pack gum
bubble gum
honey
jams, jellies, & preserves
peanut butter & combin
shelf stable tomato juice
shelf stable blended veg juice
shelf stable orange juice
prune juice
shelf-stable pineapple juice
shelf-stable grapefruit juice
shelf-stable apple juice
shelf-stable grape juice
shelf-stable fruit nectars
shelf-stable blended fruit juice
shelf-stable cider
shelf-stable lemon & lime juice
shelf-stable misc fruit juice
shelf-stable misc veg juice
shelf-stable concentrates jce & drink
shelf-stable jce drinks sgl strength
pasta
canned meat stew
canned beef hash
canned poultry products
canned corn beef
canned meat spreads
canned lunch meat
canned sausage & frankfurters
canned dry beef
misc canned beef & pork
canned hams & bulk meats
canned meat dishes
evap condensed milk
powdered milk
coffee creamers
mlk & mlk-based products
pickles
relishes
peppers
ripe olives
spanish olives
appetizer relish-onion
beans/pork & beans
canned chili
pasta dishes-canned
dry packaged dinners
pizza products
Oriental food
Mexican food

instant potatoes
canned bread
canned prepared salad
misc prepared foods
mayonnaise
sandwich spreads
salad dressings-spoonable
salad dressings-pourable
spices & seasonings
salt
pepper
MSG & meat tenderizers
fruit pectin
solid shortening
cooking & salad oils
popcorn
potato chips & products
corn snacks
pretzels
snack nuts
dry toaster items
fruit rolls & bars
misc snacks dips
regular soft drinks
lo-cal soft drink
soft drink mixes
cocktail mixes:dry & bottled
breakfast drink mixes
bottled water
freezer bars
non-alcoholic wine & malt
seltzers/club soda
dehydrated soup
canned soup
bouillion
granulated sugar
brown sugar
confectioners sugar
molasses
maple syrup
corn syrup
all other syrups
instant tea & hot drink mixes
tea:packaged
tea bags
iced tea mixes & liquids
canned peas
canned wax beans
canned tomatoes
canned potatoes
canned beets
canned onions
canned butter & lima beans
canned kidney & misc beans
canned asparagus
canned spinach
canned mushrooms
canned sauerkraut
canned carrots
canned mixed vegetables
canned corn
canned green beans
tomato paste
tomato sauce
tomato puree
pumpkin
canned misc vegetables
dried rice
prepared rice
misc dried vegetables
frozen puddings
refrig & frozen toppings
frozen sweet goods
frozen pies
frozen pastry items
dinner bread & rolls-frozen
meat:frozen
frozen fish
strawberries:frozen
raspberries:frozen
melon:frozen
blueberries:frozen
mixed fruit:frozen
all other fruit frozen
orange juice:frozen
grape juice:frozen
lemonage lime-orange:frozen
fruit drinks:frozen
apple juice/apl cdr:frozen
grapefruit juice:frozen
misc frozen juices
poultry dishes:frozen
meat dishes:frozen
fish dishes:frozen
Italian dishes:frozen
Mexican dishes:frozen

Oriental dishes:frozen
other dishes:frozen
frozen potatoes
frozen pot pies
frozen mexican dinners
frozen oriental dinners
frozen hors d'ouevres
frozen dinners
frozen onion rings
frozen pizza
frozen sandwiches
frozen corn dogs
frozen & refrig seafood cocktail
misc frozen prepared foods
frozen veg/butter sauce
frozen veg/cheese sauce
frozen veg/cream sauce
frozen veg/without sauce
frozen rice/rice combo
int style frozen prep veg
american style frozen prep veg
misc frozen prep veg
frozen peas
frozen corn
frozen green & wax beans
frozen mixed veg
frozen lima beans
frozen asparagus
frozen broccoli
frozen spinach
frozen cauliflower
frozen brussel sprouts
frozen squash
frozen onions
misc frozen veg
frozen veg deluxe
frozen veg southern
frozen waffles
frozen french toast & pancakes
frozen & refrig breakfasts
frozen & refrig nondairy creamers
frozen quiche & pancake batter
frozen microwave popcorn
other frozen food
natural cheese
processed cheese
cream cheese
refrig yogurt
frozen yogurt
misc yogurt products
frankfurters
lunch meat
bacon
frozen & refrig breakfast sausage
frozen & refrig dinner sausage
butter
butter/marg blends
margarine
refrig cookies
refrig dinner rolls
refrig pastries
refrig biscuits
refrig & frozen english muffins
refrig bread dough
refrig pie crust
misc refrig dough products
refrig salad/gelatin/parfait
refrig horseradish
refrig salad dressing & sauce
refrig herring
refrig yeast
refrig orange juice
refrig puddings
refrig grapefruit juice
refrig apple juice/cider
misc refrig juices
refrig drinks
refrig mexican foods
refrig dips
misc refrig items

REFERENCES

Byrne, P.J., Capps, Jr., O., & Saha, A. (1996). Analysis of food away from home expenditure patterns for U.S. households. American Journal of Agricultural Economics, 78, 614–627.

Connor, J.M. (1981). Food product proliferation: a market structure analysis. American Journal of Agricultural Economics, 63, 607–617.

Food Marketing Review, 1994–95 Economic Research Service, U.S. Department of Agriculture, Agricultural Economic Report Number 743, Sept. 1996.

Gallo, A.E. (1999). Fewer food products introduced in last 3 years. FoodReview, September–December, 27–29.
Hiemstra, S.J., & Kim, W.G. (1995). Factors affecting expenditures for food away from home in commercial establishments by type of eating place and meal occasion. Hospitality Research Journal, 19(3):15–31.
"How grocers are fighting giant rivals." (1997, March 27). The Wall Street Journal. Vol. 230, No. 102, Pg. B-11.
Jackson, L.F. (1984). Hierarchic demand and the Engel Curve for variety. Review of Economics and Statistics, 66, 8–15.
Jones, E., & Mustiful, B.W. (1996). Purchasing behavior of higher- and lower-income shoppers: a look at breakfast cereals. Applied Economics, 28, 131–137.
Lee, J. (1987). The demand for a varied diet with econometric models for count data. American Journal of Agricultural Economics, 69, 687–692.
Lee, J., & Brown, M.G. (1989). Consumer demand for food diversity. Southern Journal of Agricultural Economics, 21, 47–53.
McCracken, V.A., & Brandt, J.A. (1987). Household consumption of food-away-from-home: total expenditures by type of food facility. American Journal of Agricultural Economics, 69, 274–284.
Messinger, P.R., & Narasimham (1997). "A Model of Retail Formats Based on Consumers' Economizing on Shopping Time" Marketing Science 16(1) 1–23.
Nayga, R.M., Jr. (1998). A sample selection model for prepared food expenditures. Applied Economics, 30, 345–352.
Patil, G.P., & Taillie, C. (1982). Diversity as a concept and its measurement. Journal of the American Statistical Association, 77, 548–561.
Park, J.L., & Capps, Jr., O. (1997). The demand for prepared meals by U.S. households. American Journal of Agricultural Economics, 79, 814–824.
Progressive Grocer. (1987). 54th annual report of the grocery industry. Stamford, CT: MacLean Hunter Media, Inc.
Progressive Grocer. (1991). Marketing guidebook. Stamford, CT: MacLean Hunter Media, Inc.
Scherer, F.M. (1979). The welfare economics of product variety: an application to the ready-to-eat cereals industry. Journal of Industrial Economics, 28, 113–134.
Shonkwiler, J.S., Lee, J., & Taylor, T.G. (1987). An empirical model of the demand for a varied diet. Applied Economics, 19, 1403–1410.
Silberberg, E. (1985). Nutrition and the demand for tastes. Journal of Political Economy, 93, 881–900.
Stigler, G. (1945). The cost of subsistence. Journal of Farm Economics, 27, 303–314.
Theil, H., & Finke, R. (1983). The consumer's demand for diversity. European Economic Review, 23, 348–359.
Yen, S.T. (1993). Working wives and food away from home: the box-cox double hurdle model. American Journal of Agricultural Economics, 75, 884–895.

Mark D. Jekanowski *is an Agricultural Economist in the Food Markets branch of the Economic Research Service, USDA. He received his Ph.D. from Purdue University in 1998, MS from Purdue University in 1993, and BS from University of Massachusetts in 1991. His current research interests include industry structure and firm conduct in the retail grocery and food-away-from-home markets.*

James K. Binkley *is a Professor of Agricultural Economics at Purdue University. He received his Ph.D. from Virginia Polytechnic Institute in 1977, MA from Washington State University in 1972, and BA from Indiana University in 1968. His current research interests include retail food markets, nutrition demand, and relation between food consumption and health outcomes.*